OECD GUIDELINE FOR THE TESTING OF CHEMICALS

DRAFT PROPOSAL FOR A NEW TEST GUIDELINE

Skin Sensitisation: Local Lymph Node Assay: DA Version

INTRODUCTION

1. OECD Guidelines for the Testing of Chemicals are periodically reviewed in light of scientific progress, changing regulatory needs, and animal welfare considerations. The first Test Guideline (TG) for the determination of skin sensitisation in the mouse, the radiolabelled Local Lymph Node Assay (LLNA; TG 429) was adopted in 2002 (1). The details of the validation of the LLNA and a review of the associated work have been published (2)(3)(4)(5)(6). A modified LLNA test method utilising non-radiolabelled components and the measurement of adenosine triphosphate (ATP) content by bio-luminescence as the endpoint has been developed by Daicel Chemical Industries, Ltd., the LLNA: DA test method. The LLNA: DA test method has been validated and based on a formal peer review, the LLNA: DA is considered useful for identifying skin sensitising and non-sensitising test substances, with certain limitations (7)(8)(9)(10). This is the third TG to be promulgated for assessing skin sensitisation potential of chemicals in animals. Test Guideline 406 utilises guinea pig tests, notably the guinea pig maximisation test and the Buehler test (11). The LLNA: DA provides certain advantages over TG 406 with regard to both scientific progress and animal welfare.

2. Similar to the LLNA, the LLNA: DA studies the induction phase of skin sensitisation and provides quantitative data suitable for dose response assessment. Furthermore, an ability to detect skin sensitizers without the necessity for using a radiolabel for DNA eliminates the potential for occupational exposure to radioactivity and waste disposal issues. This in turn may allow for the increased use of mice to detect skin sensitizers, which could further reduce the use of guinea pigs to test for skin sensitisation potential (*i.e.* TG 406) (11). A reduced LLNA: DA (rLLNA: DA) protocol that uses fewer animals is also described in this TG (1)(13)(14). The rLLNA: DA may be used for the hazard classification of skin sensitising test substances when dose-response information is not needed provided there is adherence to all other LLNA: DA protocol specifications as described in this TG. The rLLNA: DA should not be used for the hazard identification of skin sensitising test substances when dose-response information is needed.

DEFINITIONS

3. Definitions used are provided in Annex 1.

INITIAL CONSIDERATIONS AND LIMITATIONS

4. The LLNA: DA is a modified LLNA method for identifying potential skin sensitising test substances, with specific limitations. This does not necessarily imply that in all instances the LLNA: DA should be used in place of the LLNA or guinea pig tests (*i.e.* TG

406) (11), but rather that the assay may be employed as an alternative in which positive and negative results generally no longer require further confirmation (7)(8). The testing laboratory should consider all available information on the test substance prior to conducting the study. Such information will include the identity and chemical structure of the test substance; its physicochemical properties; the results of any other *in vitro* or *in vivo* toxicity tests on the test substance; and toxicological data on structurally related test substances.

5. The LLNA: DA is an *in vivo* method and, as a consequence, will not eliminate the use of animals in the assessment of allergic contact sensitising activity. It has, however, the potential to reduce the number of animals required for this purpose (*e.g.* reducing the number of guinea pigs used when the LLNA: DA is used instead of guinea pig assays or the LLNA where the use of radioactivity is discouraged). Moreover, the LLNA: DA offers a substantial refinement of the way in which animals are used for allergic contact sensitisation testing. The LLNA: DA is based upon consideration of immunological events stimulated by chemicals during the induction phase of sensitisation. Unlike guinea pig tests (*i.e.* TG 406) (11), the LLNA: DA does not require that challenge-induced dermal hypersensitivity reactions be elicited. Thus, the LLNA: DA reduces animal distress. Despite the advantages of the LLNA: DA over TG 406 (11), there are certain limitations that may necessitate the use of TG 406 (11) (*e.g.* the testing of certain metals, false positive findings with certain skin irritants [such as some surfactant-type materials] (6)(15), solubility of the test material, or test substance classes or materials containing functional groups shown to act as potential confounders (16)). Limitations that have been identified for the LLNA have been recommended to apply also to the LLNA: DA (7). For the validation database of 44 substances, the LLNA: DA correctly identified all 32 LLNA sensitizers, but three of 12 LLNA non-sensitizers were identified as borderline positive, with Stimulation Index (SI) values between 1.8 and 2.5 (7). Additionally, the use of the LLNA: DA might not be appropriate for testing test substances that affect ATP levels (*e.g.* test substances that function as ATP inhibitors) or those that affect the accurate measurement of intracellular ATP (*e.g.* presence of ATP degrading enzymes, presence of extracellular ATP in the lymph node). Other than such identified limitations and considerations, the LLNA: DA should be applicable for testing any test substances unless there are properties associated with these materials that may interfere with the accuracy of the LLNA: DA.

PRINCIPLE OF THE TEST

6. The basic principle underlying the LLNA: DA is that sensitizers induce proliferation of lymphocytes in the lymph nodes draining the site of test substance application. This proliferation is proportional to the dose and to the potency of the applied allergen and provides a simple means of obtaining a quantitative measurement of sensitisation. Proliferation is measured by comparing the mean proliferation in each test group to the mean proliferation in the vehicle treated control (VC) group. The ratio of the mean proliferation in each treated group to that in the concurrent VC group, termed the SI, is determined, and should be $\geq$1.8 before further evaluation of the test substance as a potential skin sensitizer is warranted. The methods described here are based on the use of measuring ATP content by bioluminescence (known to correlate with living cell number) (17) to indicate an increased number of proliferating cells in the draining auricular lymph nodes (18)(19). The

bioluminescent method utilises the luciferase enzyme to catalyse the formation of light from ATP and luciferin according to the following reaction:

$$ATP + Luciferin + O_2 \xrightarrow{Luciferase} Oxyluciferin + AMP + PP_i + CO_2 + Light$$

The emitted light intensity is linearly related to the ATP concentration and is measured using a luminometer. The luciferin-luciferase assay is a sensitive method for ATP quantitation used in a wide variety of applications (20).

DESCRIPTION OF THE ASSAY

Selection of animal species

7. The mouse is the species of choice for this test. Validation studies for the LLNA: DA were conducted exclusively with the CBA/J strain, which is therefore considered the preferred strain (9)(10). Young adult female mice, which are nulliparous and non-pregnant, are used. At the start of the study, animals should be between 8-12 weeks old, and the weight variation of the animals should be minimal and not exceed 20% of the mean weight. Alternatively, other strains and males may be used when sufficient data are generated to demonstrate that significant strain and/or gender-specific differences in the LLNA: DA response do not exist.

Housing and feeding conditions

8. Mice should be group housed (21), unless adequate scientific rationale for housing mice individually is provided. The temperature of the experimental animal room should be 22ºC (±3ºC). Although the relative humidity should be at least 30% and preferably not exceed 70%, other than during room cleaning, the aim should be 50-60%. Lighting should be artificial, the sequence being 12 hours light, 12 hours dark. For feeding, conventional laboratory diets may be used with an unlimited supply of drinking water.

Preparation of animals

9. The animals are randomly selected, marked to permit individual identification (but not by any form of ear marking), and kept in their cages for at least five days prior to the start of dosing to allow for acclimatisation to the laboratory conditions. Prior to the start of treatment all animals are examined to ensure that they have no observable skin lesions.

Preparation of dosing solutions

10. Solid test substances should be dissolved or suspended in solvents/vehicles and diluted, if appropriate, prior to application to an ear of the mice. Liquid test substances may be applied neat or diluted prior to dosing. Insoluble materials, such as those generally seen in medical devices, should be subjected to an exaggerated extraction in an appropriate solvent to reveal all extractable constituents for testing prior to application to an ear of the mice. Test substances should be prepared daily unless stability data demonstrate the acceptability of storage.

Reliability check

11. Positive controls (PC) are used to demonstrate appropriate performance of the assay by responding with adequate and reproducible sensitivity to a sensitising test substance for which the magnitude of the response is well characterised. Inclusion of a concurrent PC is recommended because it demonstrates competency of the laboratory to successfully conduct each assay and allows for an assessment of intra- and inter-laboratory reproducibility and comparability. A PC for each study is also required by some regulatory authorities. Accordingly, the routine use of a concurrent PC is encouraged to avoid the need for additional animal testing to meet such requirements that might arise from the use of a periodic PC (see paragraph 12). The PC should produce a positive LLNA: DA response at an exposure level expected to give an increase in the SI $\geq$ 1.8 over the negative control (NC) group. The PC dose should be chosen such that the induction is reproducible but not excessive (*e.g.* SI > 10 would be considered excessive). Preferred PC test substances are 25% hexyl cinnamic aldehyde (Chemical Abstracts Service [CAS] No 101-86-0) and eugenol (CAS No 97-53-0) in acetone: olive oil. There may be circumstances in which, given adequate justification, other PC test substances, meeting the above criteria, may be used.

12. While inclusion of a concurrent PC group is recommended, there may be situations in which periodic testing (*i.e.* at intervals $\leq$6 months) of the PC test substance may be adequate for laboratories that conduct the LLNA: DA regularly (*i.e.* conduct the LLNA: DA at a frequency of no less than once per month) and have an established historical PC database that demonstrates the laboratory's ability to obtain reproducible and accurate results with PCs. Adequate proficiency with the LLNA: DA can be successfully demonstrated by generating consistent positive results with the PC in at least 10 independent tests conducted within a reasonable period of time (*i.e.* less than one year).

13. A concurrent PC group should always be included when there is a procedural change to the LLNA: DA (*e.g.* change in trained personnel, change in test method materials and/or reagents, change in test method equipment, change in source of test animals), and such changes should be documented in laboratory reports. Consideration should be given to the impact of these changes on the adequacy of the previously established historical database in determining the necessity for establishing a new historical database to document consistency in the PC results.

14. Investigators should be aware that the decision to conduct a PC on a periodic basis instead of concurrently has ramifications on the adequacy and acceptability of negative study results generated without a concurrent PC during the interval between each periodic PC study. For example, if a false negative result is obtained in the periodic PC study, negative test substance results obtained in the interval between the last acceptable periodic PC study and the unacceptable periodic PC study may be questioned. Implications of these outcomes should be carefully considered when determining whether to include concurrent PCs or to only conduct periodic PCs. Consideration should also be given to using fewer animals in the concurrent PC group when this is scientifically justified and if the laboratory demonstrates, based on laboratory-specific historical data, that fewer mice can be used (22).

15. Although the PC test substance should be tested in the vehicle that is known to elicit a consistent response (*e.g.* acetone: olive oil), there may be certain regulatory situations in which testing in a non-standard vehicle (clinically/chemically relevant formulation) will also be necessary (1). If the concurrent PC test substance is tested in a different vehicle than the test substance, then a separate vehicle control for the concurrent PC should be included.

16. In instances where test substances of a specific chemical class or range of responses are being evaluated, benchmark test substances may also be useful to demonstrate that the test method is functioning properly for detecting the skin sensitisation potential of these types of test substances. Appropriate benchmark test substances should have the following properties:

- structural and functional similarity to the class of the test substance being tested;
- known physical/chemical characteristics;
- supporting data from the LLNA: DA;
- supporting data from other animal models and/or from humans.

TEST PROCEDURE

Number of animals and dose levels

17. A minimum of four animals is used per dose group, with a minimum of three concentrations of the test substance, plus a concurrent NC group treated only with the vehicle for the test substance, and a PC (concurrent or recent, based on laboratory policy in considering see paragraphs 11-15). Except for absence of treatment with the test substance, animals in the control groups should be handled and treated in a manner identical to that of animals in the treatment groups.

18. Dose and vehicle selection should be based on the recommendations given in references (2) and (24). Consecutive doses are normally selected from an appropriate concentration series such as 100%, 50%, 25%, 10%, 5%, 2.5%, 1%, 0.5%, etc. Adequate scientific rationale should accompany the selection of the concentration series used. All existing toxicological information (*e.g.* acute toxicity and dermal irritation) and structural and physicochemical information on the test substance of interest (and/or structurally related test substances) should be considered, where available, in selecting the three consecutive concentrations so that the highest concentration maximises exposure while avoiding systemic toxicity and/or excessive local skin irritation (24)(25). In the absence of such information, an initial pre-screen test may be necessary (see paragraphs 21-23).

19. The vehicle should not interfere with or bias the test result and should be selected on the basis of maximising the solubility in order to obtain the highest concentration achievable while producing a solution/suspension suitable for application of the test substance. Recommended vehicles are acetone: olive oil (4:1 v/v), *N*,*N*-dimethylformamide, methyl ethyl ketone, propylene glycol, and dimethyl sulphoxide (6) but others may be used if sufficient scientific rationale is provided. In certain situations it may be necessary to use a

clinically relevant solvent or the commercial formulation in which the test substance is marketed as an additional control. Particular care should be taken to ensure that hydrophilic materials are incorporated into a vehicle system, which wets the skin and does not immediately run off by incorporation of appropriate solubilisers (*e.g.* 1% Pluronic® L92). Thus, wholly aqueous vehicles are to be avoided.

20. The processing of lymph nodes from individual mice allows for the assessment of inter-animal variability and a statistical comparison of the difference between test substance and vehicle control group measurements (see paragraph 34). In addition, evaluating the possibility of reducing the number of mice in the PC group is only feasible when individual animal data are collected (22). Further, some national regulatory authorities require the collection of individual animal data. Regular collection of individual animal data provides an animal welfare advantage by avoiding duplicate testing that would be necessary if the test substance results originally collected in one manner (*e.g.* via pooled animal data) were to be considered later by regulatory authorities with other requirements (*e.g.* individual animal data).

Pre-screen test

21. In the absence of information to determine the highest dose to be tested (see paragraph 18), a pre-screen test should be performed in order to define the appropriate dose level to test in the LLNA: DA. The purpose of the pre-screen test is to provide guidance for selecting the maximum dose level to use in the main LLNA: DA study, where information on the concentration that induces systemic toxicity (see paragraph 23) and/or excessive local skin irritation (see paragraph 23) is not available. The maximum dose level tested should be 100% of the test substance for liquids or the maximum possible concentration for solids or suspensions, unless available information suggests that this concentration induces systemic toxicity and/or excessive local irritation after topical application in the mouse.

22. The pre-screen test is conducted under conditions identical to the main LLNA: DA study, except there is no assessment of lymph node proliferation and fewer animals per dose group can be used. One or two animals per dose group are suggested. All mice will be observed daily for any clinical signs of systemic toxicity or local irritation at the application site. Body weights are recorded pre-test and prior to termination (Day 8). Both ears of each mouse are observed for erythema and scored using Table 1 (25). Ear thickness measurements are taken using a thickness gauge (*e.g.* digital micrometer or Peacock Dial thickness gauge) on Day 1 (pre-dose), Day 3 (approximately 48 hours after the first dose), Day 7 (24 hours prior to termination) and Day 8. Additionally on Day 8, ear thickness could be determined by ear punch weight determinations. Excessive local irritation is indicated by an erythema score ≥ 3 and/or ear thickness of $\geq 25\%$ on any day of measurement (26)(27). The highest dose selected for the main LLNA: DA study will be the next lower dose in the pre-screen concentration series (see paragraph 18) that does not induce systemic toxicity and/or excessive local skin irritation.

Table 1. Erythema scores

Observation	Score
No erythema	0
Very slight erythema (barely perceptible)	1
Well-defined erythema	2
Moderate to severe erythema	3
Severe erythema (beet redness) to eschar formation preventing grading of erythema	4

23. In addition to a 25% increase in ear thickness (26)(27), a statistically significant increase in ear thickness in the treated mice compared to control mice has also been used to identify irritants in the LLNA (28)(29)(30)(31)(32)(33)(34). However, while statistically significant increases can occur when ear thickness is less than 25% they have not been associated specifically with excessive irritation (30)(32)(33)(34).

24. The following clinical observations may indicate systemic toxicity (1) when used as part of an integrated assessment and therefore may indicate the maximum dose level to use in the main LLNA: changes in nervous system function (*e.g.* pilo-erection, ataxia, tremors, and convulsions); changes in behaviour (*e.g.* aggressiveness, change in grooming activity, marked change in activity level); changes in respiratory patterns (*i.e.* changes in frequency and intensity of breathing such as dyspnea, gasping, and rales), and changes in food and water consumption. In addition, signs of lethargy and/or unresponsiveness and any clinical signs of more than slight or momentary pain and distress, or a >5% reduction in body weight from Day 1 to Day 6 and mortality should be considered in the evaluation.

Reduced LLNA

25. Use of a rLLNA: DA protocol instead of the multi-dose LLNA: DA has the potential to reduce the number of animals used in a test by omitting the middle and low dose groups (1)(13)(14). The reduction in number of dose groups is the only difference between the LLNA: DA and the rLLNA: DA test method protocols and for this reason the rLLNA: DA does not provide dose-response information. Therefore, the rLLNA: DA should not be used when dose-response information is needed. Like the multi-dose LLNA: DA, the test substance concentration evaluated in the rLLNA: DA should be the maximum concentration that does not induce overt systemic toxicity and/or excessive local skin irritation in the mouse (see paragraph 18).

Main study experimental schedule

26. The experimental schedule of the assay is as follows:

- *Day 1:*
  Individually identify and record the weight of each animal and any clinical observations. Apply 1% sodium lauryl sulphate (SLS) aqueous solution to the dorsum of each ear by using a brush dipped in the SLS solution to cover the

entire dorsum of each ear with four to five strokes. One hour after the SLS treatment, apply 25 µL of the appropriate dilution of the test substance, the vehicle alone, or the concurrent PC (see paragraphs 11-15), to the dorsum of each ear.

- *Days 2, 3 and 7:*
  Repeat the 1% SLS aqueous solution pre-treatment and test substance application procedure carried out on Day 1.
- *Days 4, 5, and 6:*
  No treatment.
- *Day 8:*
  Record the weight of each animal and any clinical observations. Approximately 24 to 30 hours after the start of application on Day 7 humanely kill the animals. Excise the draining auricular lymph nodes from each mouse ear and process separately in phosphate buffered saline (PBS) for each animal. Details and diagrams of the node identification and dissection can be found in reference (22). To further monitor the local skin response in the main study, additional parameters such as scoring of ear erythema or ear thickness measurements (obtained either by using a thickness gauge, or ear punch weight determinations at necropsy) may be included in the study protocol.

Preparation of cell suspensions

27. From each mouse, a single-cell suspension of lymph node cells excised bilaterally is prepared by sandwiching the lymph nodes between two glass slides and applying light pressure to crush the nodes. After confirming that the tissue has spread out thinly pull the two slides apart. Suspend the tissue on both slides in PBS by holding each slide at an angle over the Petri dish and rinsing with PBS while concurrently scraping the tissue off of the slide with a cell scraper. A total volume of 1 mL PBS should be used for rinsing both slides. The tissue suspension in the Petri dish should be homogenised lightly with the cell scraper. A 20 µL aliquot of the homogenised suspension is then collected with a micropipette, taking care not to take up the membrane that is visible to the eye, and subsequently mixed with 1.98 mL of PBS to yield a 2 mL sample. A second 2 mL sample is then prepared using the same procedure so that two samples are prepared for each animal. These procedures should be performed while wearing gloves and a mask and micropipette tips should be sterile.

Determination of cellular proliferation (measurement of ATP content of lymphocytes)

28. Increases in ATP content in the lymph nodes are measured by the luciferin/luciferase method using an ATP measurement kit, which measures bioluminescence in relative luminescence units (RLU). The assay time from time of animal sacrifice to measurement of ATP content for each individual animal should be kept uniform, within approximately 30 minutes, because the ATP content is considered to gradually decrease with time after animal sacrifice (9). Thus, the series of procedures from excision of auricular lymph nodes to ATP measurement should be completed within 20 minutes by the pre-determined time schedule that is the same for each animal. ATP luminescence should be measured in each 2 mL sample so that a total of two ATP measurements are collected for

each animal. The mean ATP luminescence is then determined and used in subsequent calculations (see paragraph 32).

OBSERVATIONS

Clinical observations

29. Each mouse should be carefully observed at least once daily for any clinical signs, either of local irritation at the application site or of systemic toxicity. All observations are systematically recorded with records being maintained for each mouse. Monitoring plans should include criteria to promptly identify those mice exhibiting systemic toxicity, excessive local skin irritation, or corrosion of skin for euthanasia.

Body weights

30. As stated in paragraph 26, individual animal body weights should be measured at the start of the test and at the scheduled kill.

CALCULATION OF RESULTS

31. Results for each treatment group are expressed as the mean SI. The SI is derived by dividing the mean RLU/mouse within each test substance group and the concurrent PC group by the mean RLU/mouse for the solvent/vehicle control group. The average SI for vehicle treated controls is then one.

32. The decision process regards a result as positive when $SI \geq 1.8$ (7). However, the strength of the dose-response, the statistical significance and the consistency of the solvent/vehicle and positive control responses may also be used when determining whether a borderline result is declared positive (2)(3)(1).

33. For a borderline positive response between an SI of 1.8 and 2.5, users may want to consider additional information such as dose response, evidence of systemic toxicity or excessive irritation, and where appropriate, statistical significance together with SI values to confirm that such results are positives (7). Consideration should also be given to various properties of the test substance, including whether it has a structural relationship to known skin sensitizers, whether it causes excessive skin irritation in the mouse, and the nature of the dose response seen. These and other considerations are discussed in detail elsewhere (4).

34. Collecting data at the level of the individual mouse will enable a statistical analysis for presence and degree of dose response in the data. Any statistical assessment could include suitably adjusted comparisons of test groups (*e.g.* pair-wise dosed group versus concurrent solvent/vehicle control comparisons). Statistical analyses may include, for instance, linear regression or William's test to assess dose-response trends, and Dunnett's test for pair-wise comparisons. In choosing an appropriate method of statistical analysis, the investigator should maintain an awareness of possible inequalities of variances and other related problems that may necessitate a data transformation or a non-parametric statistical analysis. In any case, the investigator may need to carry out SI calculations and statistical analyses with and without certain data points (sometimes called "outliers").

DATA AND REPORTING

Data

35. Data should be summarised in tabular form showing the individual animal RLU values, the group mean RLU/animal, its associated error term (*e.g.* SD, SEM), and the mean SI for each dose group compared against the concurrent solvent/vehicle control group.

Test report

36. The test report should contain the following information:

Test substance and control test substance:

- identification data (*e.g.* CAS number, if available; source; purity; known impurities; lot number);
- physical nature and physicochemical properties (*e.g.* volatility, stability, solubility);
- if mixture, composition and relative percentages of components.

Solvent/vehicle:

- identification data (purity; concentration, where appropriate; volume used);
- justification for choice of vehicle.

Test animals:

- source of CBA mice;
- microbiological status of the animals, when known;
- number and age of animals;
- source of animals, housing conditions, diet, etc.

Test conditions:

- details of test substance preparation and application;
- justification for dose selection (including results from pre-screen test, if conducted);
- vehicle and test substance concentrations used, and total amount of test substance applied;
- details of food and water quality (including diet type/source, water source);
- details of treatment and sampling schedules;
- methods for measurement of toxicity;
- criteria for considering studies as positive or negative;
- details of any protocol deviations and an explanation on how the deviation affects the study design and results.

Reliability check:

- a summary of results of latest reliability check, including information on test substance, concentration and vehicle used;
- concurrent and/or historical PC and concurrent negative (solvent/vehicle) control data for testing laboratory;
- if a concurrent PC was not included, the date and laboratory report for the most recent periodic PC and a report detailing the historical PC data for the laboratory justifying the basis for not conducting a concurrent PC.

Results:

- individual weights of mice at start of dosing and at scheduled kill; as well as mean and associated error term (*e.g.* SD, SEM) for each treatment group;
- time course of onset and signs of toxicity, including dermal irritation at site of administration, if any, for each animal;
- time of animal sacrifice and time of ATP measurement for each animal;
- a table of individual mouse RLU values and SI values for each dose treatment group;
- mean and associated error term (*e.g.* SD, SEM) for RLU/mouse for each treatment group and the results of outlier analysis for each treatment group;
- calculated SI and an appropriate measure of variability that takes into account the inter-animal variability in both the test substance and control groups;
- dose response relationship;
- statistical analyses, where appropriate.

Discussion of results:

- a brief commentary on the results, the dose-response analysis, and statistical analyses, where appropriate, with a conclusion as to whether the test substance should be considered a skin sensitizer.

406 **LITERATURE**

407 (1) OECD (2002), Skin Sensitisation: Local Lymph Node Assay. OECD Guideline for the
408 Testing of Chemicals No 429, Paris, France. Available at:
409 [http://www.oecd.org/env/testguidelines].

410 (2) Chamberlain, M. and Basketter, D.A. (1996), The local lymph node assay: status of
411 validation. *Food Chem, Toxicol*,, 34, 999-1002.

412 (3) Basketter, D.A., Gerberick, G.F., Kimber, I. and Loveless, S.E. (1996), The local
413 lymph node assay: A viable alternative to currently accepted skin sensitisation tests.
414 *Food Chem, Toxicol*,, 34, 985-997.

415 (4) Basketter, D.A., Gerberick, G.F. and Kimber, I. (1998), Strategies for identifying
416 false positive responses in predictive sensitisation tests. *Food Chem. Toxicol.*, 36,
417 327-333.

418 (5) Van Och, F.M.M., Slob, W., De Jong, W.H., Vandebriel, R.J. and Van Loveren, H.
419 (2000), A quantitative method for assessing the sensitising potency of low
420 molecular weight chemicals using a local lymph node assay: employment of a
421 regression method that includes determination of uncertainty margins. *Toxicol*., 146,
422 49-59.

423 (6) ICCVAM (1999), The murine local lymph node Assay: A test method for assessing
424 the allergic contact dermatitis potential of chemicals/compounds: The results of an
425 independent peer review evaluation coordinated by the Interagency Coordinating
426 Committee on the Validation of Alternative Methods (ICCVAM) and the National
427 Toxicology Program Center for the Evaluation of Alternative Toxicological
428 Methods (NICETAM). NIH Publication No: 99-4494. Research Triangle Park, N.C.
429 Available at: [http://iccvam.niehs.nih.gov/docs/immunotox_docs/llna/llnarep.pdf].

430 (7) ICCVAM (2009), ICCVAM Test Method Evaluation Report. Nonradioactive local
431 lymph node assay: modified by Daicel Chemical Industries, Ltd., based on ATP
432 content test method protocol (LLNA: DA). NIH Publication No. XX-XXXX.
433 Research Triangle Park, NC: National Institute of Environmental Health Sciences.
434 Available at: [XXX].

435 (8) ICCVAM (2009), Independent Scientific Peer Review Panel Report: Updated
436 validation status of new versions and applications of the murine local lymph node
437 assay: a test method for assessing the allergic contact dermatitis potential of
438 chemicals and products. Research Triangle Park, NC: National Institute of
439 Environmental Health Sciences. Available at:
440 [http://iccvam.niehs.nih.gov/docs/immunotox_docs/LLNAPRPRept2009.pdf].

441 (9) Idehara, K., Yamagishi, G., Yamashita, K. and Ito, M. (2008), Characterization and
442 evaluation of a modified local lymph node assay using ATP content as a non-radio
443 isotopic endpoint. *J. Pharmacol. Toxicol. Meth.*, 58, 1-10.

444 (10) Omori, T., Idehara, K., Kojima, H., Sozu, T., Arima, K., Goto, H., Hanada, T.,
445 Ikarashi, Y., Inoda, T., Kanazawa, Y., Kosaka, T., Maki, E., Morimoto, T., Shinoda,
446 S., Shinoda, N., Takeyoshi, M., Tanaka, M., Uratani, M., Usami, M., Yamanaka, A.,
447 Yoneda, T., Yoshimura, I. and Yuasa, A. (2008), Interlaboratory validation of the
448 modified murine local lymph node assay based on adenosine triphosphate
449 measurement. *J. Pharmacol. Toxicol. Meth.*, 58, 11-26.

(11) OECD (1992), Skin Sensitisation. OECD Guideline for Testing of Chemicals No 406, Paris, France. Available at: [http://www.oecd.org/env/testguidelines].

(12) Kimber, I., Dearman, R.J., Betts, C.J., Gerberick, G.F., Ryan, C.A., Kern, P.S., Patlewicz, G.Y. and Basketter, D.A. (2006), The local lymph node assay and skin sensitization: a cut-down screen to reduce animal requirements? *Contact Dermatitis*, 54, 181-185.

(13) ESAC (2007), Statement on the Reduced Local Lymph Node Assay (rLLNA). European Commission Directorate General, Joint Research Centre, Institute for Health and Consumer Protection, European Centre for the Validation of Alternative Methods. April 2007. Available: [http://ecvam.jrc.it/ft_doc/ESAC26_statement_rLLNA_20070525-1.pdf].

(14) ICCVAM (2009), The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) Test Method Evaluation Report. The Reduced Murine Local Lymph Node Assay: An Alternative Test Method Using Fewer Animals to Assess the Allergic Contact Dermatitis Potential of Chemicals and Products. NIH Publication Number 09-6439. Research Triangle Park, NC: National Institute of Environmental Health Sciences.

(15) Kreiling, R., Hollnagel, H.M., Hareng, L., Eigler, L., Lee, M.S., Griem, P., Dreessen, B., Kleber, M., Albrecht, A., Garcia, C. and Wendel, A. (2008), Comparison of the skin sensitizing potential of unsaturated compounds as assessed by the murine local lymph node assay (LLNA) and the guinea pig maximization test (GPMT). *Food Chem. Toxicol*., 46, 1896-1904.

(16) Basketter, D., Ball, N., Cagen, S., Carrilo, J.C., Certa, H., Eigler, D., Garcia, C., Esch, H., Graham, C., Haux, C., Kreiling, R. and Mehling, A. (2009), Application of a weight of evidence approach to assessing discordant sensitisation datasets: implications for REACH. *Reg. Toxicol. Pharmacol*., 55, 90-96.

(17) Crouch, S.P., Kozlowski, R., Slater, K.J. and Fletcher J. (1993). The use of ATP bioluminescence as a measure of cell proliferation and cytotoxicity. *J. Immunol. Meth*., 160, 81-88.

(18) Ishizaka, A., Tono-oka, T. and Matsumoto, S. (1984), Evaluation of the proliferative response of lymphocytes by measurement of intracellular ATP. J. Immunol. Meth., 72, 127-132.

(19) Dexter, S.J., Cámara, M., Davies, M. and Shakesheff, K.M. (2003), Development of a bioluminescent ATP assay to quantify mammalian and bacterial cell number from a mixed population. *Biomat.*, 24, 27-34.

(20) Lundin A. (2000), Use of firefly luciferase in ATP-related assays of biomass, enzymes, and metabolites. *Meth. Enzymol*., 305, 346-370.

(21) ILAR (1996), Institute of Laboratory Animal Research (ILAR) Guide for the Care and Use of Laboratory Animals. 7th ed. Washington, DC: National Academies Press.

(22) ICCVAM (2009), Recommended Performance Standards: Murine Local Lymph Node Assay. NIH Publication Number 09-7357, Research Triangle Park, NC: National Institute of Environmental Health Sciences. Available at: [http://iccvam.niehs.nih.gov/docs/immunotox_docs/llna-ps/LLNAPerfStds.pdf].

(23) McGarry, H.F. (2007), The murine local lymph node assay: regulatory and potency considerations under REACH. *Toxicol*., 238, 71-89.

(24) Kimber, I., Dearman, R.J., Scholes E.W. and Basketter, D.A. (1994), The local lymph node assay: developments and applications. *Toxicol*., 93, 13-31.

(25) OECD (2002), Acute Dermal Irritation/Corrosion. OECD Guideline for Testing of Chemicals No.404, Paris: Available at: [http://www.oecd.org/env/testguidelines].

(26) Reeder, M.K., Broomhead, Y.L., DiDonato, L. and DeGeorge, G.L. (2007), Use of an enhanced local lymph node assay to correctly classify irritants and false positive substances. *Toxicologist*, 96 (S-1), 235.

(27) ICCVAM (2009), Nonradioactive Murine Local Lymph Node Assay: Flow Cytometry Test Method Protocol (LLNA: BrdU-FC) Revised Draft Background Review Document. Research Triangle Park, NC: National Institute of Environmental Health Sciences. Available at: [http://iccvam.niehs.nih.gov/methods/immunotox/fcLLNA/BRDcomplete.pdf].

(28) Hayes, B.B., Gerber, P.C., Griffey, S.S. and Meade, B.J. (1998), Contact hypersensitivity to dicyclohexylcarbodiimide and diisopropylcarbodiimide in female B6C3F1 mice. *Drug. Chem. Toxicol*., 21, 195-206.

(29) Homey, B., von Schilling, C., Blumel, J., Schuppe, H.C., Ruzicka, T., Ahr, H.J., Lehmann, P. and Vohr, V.W. (1998), An integrated model for the differentiation of chemical-induced allergic and irritant skin reactions. *Toxicol. Appl. Pharmacol.*, 153, 83-94.

(30) Woolhiser, M.R., Hayes, B.B. and Meade, B.J. (1998), A combined murine local lymph node and irritancy assay to predict sensitization and irritancy potential of chemicals. *Toxicol. Meth.*, 8, 245-256.

(31) Hayes, B.B. and Meade, B.J. (1999), Contact sensitivity to selected acrylate compounds in B6C3F1 mice: relative potency, cross reactivity, and comparison of test methods. *Drug Chem. Toxicol*., 22, 491-506.

(32) Ehling, G., Hecht, M., Heusener, A., Huesler, J., Gamer, A.O., van Loveren, H., Maurer, T., Riecke, K., Ullmann, L., Ulrich, P., Vandebriel, R. and Vohr, H.W. (2005), A European inter-laboratory validation of alternative endpoints of the murine local lymph node assay: first round. *Toxicol*., 212, 60-68.

(33) Vohr, H.W. and Ahr, H.J. (2005), The local lymph node assay being too sensitive? *Arch. Toxicol*., 79, 721-728.

(34) Patterson, R.M., Noga, E. and Germolec, D. (2007), Lack of evidence for contact sensitization by Pfiesteria extract. *Environ. Health Perspect*., 115, 1023-1028.

(35) ICCVAM (2009), Report on the ICCVAM-NICEATM/ECVAM/JaCVAM Scientific Workshop on Acute Chemical Safety Testing: Advancing In Vitro Approaches and Humane Endpoints for Systemic Toxicity Evaluations. Research Triangle Park, NC: National Institute of Environmental Health Sciences. Available at: [http://iccvam.niehs.nih.gov/methods/acutetox/Tox_workshop.htm].

(36) Kimber, I., Hilton, J., Dearman, R.J., Gerberick, G.F., Ryan, C.A., Basketter, D.A., Lea, L., House, R.V., Ladies, G.S., Loveless, S.E. and Hastings, K.L. (1998), Assessment of the skin sensitization potential of topical medicaments using the local

lymph node assay: An interlaboratory exercise. *J, Toxicol, Environ. Health,* 53 563-79.

ANNEX 1

DEFINITIONS

Accuracy: The closeness of agreement between test method results and accepted reference values. It is a measure of test method performance and one aspect of relevance. The term is often used interchangeably with "concordance" to mean the proportion of correct outcomes of a test method.

Benchmark test substance: A sensitizing or non-sensitizing substance used as a standard for comparison to a test substance. A benchmark substance should have the following properties; (i) a consistent and reliable source(s); (ii) structural and functional similarity to the class of substances being tested; (iii) known physical/chemical characteristics; (iv) supporting data on known effects, and (v) known potency in the range of the desired response.

False negative: A test substance incorrectly identified as negative or non-active by a test method, when in fact it is positive or active.

False positive: A test substance incorrectly identified as positive or active by a test, when in fact it is negative or non-active.

Hazard: The potential for an adverse health or ecological effect. The adverse effect is manifested only if there is an exposure of sufficient level.

Inter-laboratory reproducibility: A measure of the extent to which different qualified laboratories, using the same protocol and testing the same test substances, can produce qualitatively and quantitatively similar results. Inter-laboratory reproducibility is determined during the prevalidation and validation processes, and indicates the extent to which a test can be successfully transferred between laboratories, also referred to as between-laboratory reproducibility.

Intra-laboratory reproducibility: A determination of the extent that qualified people within the same laboratory can successfully replicate results using a specific protocol at different times. Also referred to as within-laboratory reproducibility.

Outlier: An outlier is an observation that is markedly different from other values in a random sample from a population.

Quality assurance: A management process by which adherence to laboratory testing standards, requirements, and record keeping procedures, and the accuracy of data transfer, are assessed by individuals who are independent from those performing the testing.

Reliability: Measures of the extent that a test method can be performed reproducibly within and between laboratories over time, when performed using the same protocol. It is assessed by calculating intra- and inter-laboratory reproducibility.

Skin sensitization: An immunological process that results when a susceptible individual is exposed topically to an inducing chemical allergen, which provokes a cutaneous immune response that can lead to the development of contact sensitization.

Stimulation Index (SI): A value calculated to assess the skin sensitization potential of a test substance that is the ratio of the proliferation in treated groups to that in the concurrent vehicle control group.

Test substance: Any material tested using this TG, whether it is a single compound or consists of multiple components (e.g. final products, formulations). When testing formulations, consideration should be given to the fact that certain regulatory authorities only require testing of the final product formulation. However, there may also be testing requirements for the active ingredient(s) of a product formulation.